CALCOLO SCIENTIFICO

17/10/2012 (1)

PROGRAMMA:

- Premesse e richiami
- SVD (decomposizione ai valori singolari)
- Metodi iterativi → Krylov (cg, gmres)
  → Multygrid (geometrica e algebrica)

Riduzione costo computazionale:

- # iterazioni ← SOLUZIONE → costo iterazione ≈ Prodotto matrice-vettore
- Precondizionamento: al posto di risolvere $Ax=b$ risolvo una cosa tipo $P^{-1}Ax=P^{-1}b$ con P adatta (ANALISI SPETTRALE)
- Trasformate veloci

1. PREMESSE e RICHIAMI

1.1 NORME

MEMO: vogliamo risolvere $Ax=b$ alla fine

def: **NORMA VETTORIALE** $\|\cdot\|: \mathbb{C}^n \to \mathbb{R}^+ \cup \{0\}$ t.c.

- $\|x\| \geq 0$ e vale $=$ sse $x=0$
- $\|\alpha x\| = |\alpha| \|x\| \quad \forall \alpha \in \mathbb{C} \ \forall x \in \mathbb{C}^n$
- $\|x+y\| \leq \|x\| + \|y\|$

ESEMPI:

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$$

$$\|x\|_1 = \left(\sum_{i=1}^{n} |x_i|\right)$$

PROP$_1$: $\|\ \|$ è una funzione uniformemente continua:

$$\forall \varepsilon > 0 \ \exists \delta = \delta(\varepsilon) : \forall x, y \in \mathbb{C}^n \quad \|x-y\| \leq \delta \implies |\|x\| - \|y\|| \leq \varepsilon$$

Questo mi dice che se misuro vettori con numeri piccoli o numeri grandi non cambia la differenza delle norme.
Detto meglio, la norma della differenza non dipende dalla grandezza dei numeri

PROP$_2$: EQUIVALENZA TOPOLOGICA: $\forall \|\|_x, \|\|_{xx} : \mathbb{C}^n \to \mathbb{R}^+ \cup \{0\}$ norme vettoriali $\exists \alpha, \beta \in \mathbb{R} \ 0 < \alpha \leq \beta$ t.c.

$$\alpha \|x\|_{xx} \leq \|x\|_x \leq \beta \|x\|_{xx} \qquad \forall x \in \mathbb{C}^n$$

Questo ci dice che alla fine possiamo usare la norma più comoda da calcolare tanto più o meno è uguale.

PROP$_3$: $\forall x \in \mathbb{C}^n$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\, \|x\|_\infty$$

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\, \|x\|_2$$

$$\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty$$

fatt Cholesky ↑ ∃ perché sdp

def: **NORMA ENERGIA**: A sdp (simm def pos / hermit.), $A = G^H G$ con G non singolare

$$\|\cdot\|_{G^HG} : \mathbb{C}^n \to \mathbb{R}^+ \cup \{0\}, \quad \|x\|_{\underbrace{G^HG}_{A}} := \|Gx\|_2 = (x^H G^H G x)^{1/2} = (x^H A x)^{1/2}$$

OSS: Siccome $A = G^HG$ posso definirla dando A o dando G ed è lo stesso

Se $A = I$ allora $\|x\|_A^2 = \|x\|_2^2 = \sum_{i=1}^{n} |x_i|^2$

Se $A = D$ diagonale def pos ($D = (d_i)$ con $d_i > 0$) $\|x\|_A^2 = x^H D x = \sum_{i=1}^{n} d_i |x_i|^2$ ↑ somma pesata

Se A è generica sdf (/hdf) posso sempre scrivere

$A = Q^H \Lambda Q$ con Q unitaria ie. $QQ^H = Q^HQ = I$ → matrice autovettori

e $\Lambda = \text{diag}(\lambda_i(A)) > 0$ ↑ autovalori

Abbiamo: $\|x\|_A^2 = x^H Q^H \Lambda Q x \underset{y = Qx}{=} y^H \Lambda y = \|y\|_\Lambda^2$

$= \sum_{i=1}^{n} \lambda_i(A) |y_i|^2$

Quindi $\|\ \|_A$ è una somma pesata, solo che dei vettori trasformati con la matrice degli autovettori.

Usare una norma energia serve per dare importanza ad alcune componenti rispetto ad altre.

A ogni norma possiamo associare un prodotto scalare:

$\|\cdot\|_2 \rightsquigarrow \langle x, y \rangle = x^H y = \sum_{i=1}^{n} \bar{x}_i y_i$

$\|\cdot\|_{G^HG} \rightsquigarrow \langle x, y \rangle_{G^HG} = \langle x, G^HGy \rangle = \langle Gx, Gy \rangle$

$(\Rightarrow \|x\|_{G^HG}^2 = \langle Gx, Gx \rangle)$

def: **NORMA MATRICIALE**: $\|\ \| : \mathbb{C}^{n \times n} \longrightarrow \mathbb{R}^+ \cup \{0\}$ t.c.

- $\|X\| \geq 0 \quad \forall X \in \mathbb{C}$ e vale $= \Leftrightarrow X = 0$
- $\|\alpha X\| = |\alpha| \|X\| \quad \forall \alpha \in \mathbb{C} \quad \forall X \in \mathbb{C}^{n \times n}$
- $\|X + Y\| \leq \|X\| + \|Y\|$
- $\|XY\| \leq \|X\| \|Y\|$ p. submoltiplicativa (non vale l'= , pensare dato due vettori $XY = 0$ ma $X, Y \neq 0$)

Come prima vogliamo le $PROP_1$ e $PROP_2$ dato due vettori dalle prime 3 proprietà.

def: $\|\cdot\|_*$ è una **NORMA MATRICIALE COMPATIBILE** con $\|\cdot\|$ norma vettoriale se:

$\|Ax\| \leq \|A\|_* \|x\| \qquad \forall x \in \mathbb{C}^n, \forall A \in \mathbb{C}^{n \times n}$

def: **NORMA MATRICIALE INDOTTA** da una norma vettoriale:

$\|A\| = \max_{\|x\| = 1} \|Ax\| = \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}$

(si verifica che è effettivamente una norma)

$PROP_4$: La norma indotta è una norma compatibile

dim: Se $x = 0$ $Ax = 0 \Rightarrow 0 = 0$ ok

Se $x \neq 0$ allora $\|Ax\| = \|x\| \frac{\|Ax\|}{\|x\|} = \|x\| \left\| A \frac{x}{\|x\|} \right\|$

poniamo $y = \frac{x}{\|x\|} \rightarrow \|y\| = 1$

$\|Ax\| = \|x\| \|Ay\| \leq \|x\| \cdot \max_{\|y\| = 1} \|Ay\| = \|x\| \|A\|$ ■

PROP$_5$: La n.m. indotta è la più piccola norma tra le norme compatibili

dim: $\|A\| = \max_{\|x\|=1} \|Ax\| \overset{\text{n.m.c.!}}{\leq} \max_{\|x\|=1} \|A\|_* \|x\| = \|A\|_*$ e cost

PROP$_6$: Indichiamo le norme indotte come le norme vettoriali, allora valgono:

$$\|A\|_1 = \max_{j=1,\dots,n} \left(\sum_{i=1}^{n} |a_{ij}| \right)$$

$$\|A\|_\infty = \max_{i=1,\dots,n} \left(\sum_{j=1}^{n} |a_{ij}| \right)$$

$$\|A\|_2 = \rho^{1/2}(A^H A)$$ dove ρ è il raggio spettrale di $A^H A$

N.B. queste non sono definizioni; le def sarebbero: $\|A\|_1 = \max_{\|x\|_1 = 1} \|Ax\|_1$ ecc...

def: **RAGGIO SPETTRALE**: $\rho(B) = \max_{i=1,\dots,n} |\lambda_i(B)|$ (max del modulo degli autovalori)

OSS: La norma 2 è proprio più brutta, perché calcolare $A^H A$ in aritmetica floating point può essere un problema, e poi anche calcolare gli autovalori, quindi conviene usare per es. $\|\ \|_\infty$ che è molto più comoda e equivalente

OSS: Se $A = A^H \Rightarrow \|A\|_1 = \|A\|_\infty$

$$\|A\|_2 = \rho^{1/2}(A^2) = \sqrt{\rho^2(A)} = \rho(A)$$

perché se $Ax = \lambda x$ allora $A(Ax) = A^2 x = \lambda^2 x$

Se A è hdp $\Rightarrow \|A\|_2 = \lambda_{max}(A)$ (posso togliere il modulo tanto sono positivi)

PROP: $\forall A \in \mathbb{C}^{n \times n}$

$$\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\, \|A\|_\infty$$

$$\frac{1}{\sqrt{n}} \|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\, \|A\|_1$$

$$\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq n \max_{i,j} |a_{ij}|$$

def: **NORMA di FROBENIUS**: $\|A\|_F := \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \left[\mathrm{tr}(A^H A) \right]^{1/2}$

N.B.: NON è una norma matriciale indotta!

dim: $\|I\|_F = \sqrt{n} > 1 = \max_{\|x\|=1} \|Ix\|$

Ogni norma indotta (si vede da qui) è t.c. $\|I\| = 1$ e quindi $\|\cdot\|_F$ non è indotta.

PROP: $A \in \mathbb{C}^{n \times n}$, $U \in \mathbb{C}^{n \times n}$ unitaria allora

$$\|UA\|_2 = \|AU\|_2 = \|UAU^H\|_2 = \|A\|_2$$

e idem per la norma $\|\cdot\|_F$.

dim: $\|B\|_2^2 = \rho(B^H B)$ $\qquad \|B\|_F^2 = \mathrm{tr}(B^H B) = \sum_{i=1}^{n} \lambda_i(B^H B)$

$(UA)^H UA = A^H U^H UA = A^H A$

$(AU)^H (AU) = U^H A^H A U \sim A^H A$ simile

$(UAU^H)^H (UAU^H) = UA^H U^H UAU^H = UA^H AU^H \sim A^H A$ simile

quindi hanno lo stesso raggio spettrale e la stessa traccia.

Conclusione: le matrici unitarie ci piacciono!

1.2 FORMA CANONICA di JORDAN

TEO: $A \in \mathbb{C}^{n\times n}$ è diagonalizzabile $A = S\Lambda S^{-1}$, $AS = \Lambda S$ sse ha n autovettori linearmente indipendenti

Non diagonalizzabile se $\mu_g < \mu_a$ ← molt. algebrica
↑ molt. geometrica

TEO: $A \in \mathbb{C}^{n\times n}$ con autovalori λ_i $i = 1, \dots, n$ con molteplicità $\mu_a(i), \mu_g(i)$

$$A \underset{\text{simile}}{\sim} J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{bmatrix}$$ diagonale a blocchi

con p = # autovalori distinti, e J_i diagonale a blocchi

$$J_i = \begin{bmatrix} C_1^{(i)} & & \\ & \ddots & \\ & & C_{\mu_g(i)}^{(i)} \end{bmatrix} \in \mathbb{C}^{\mu_a(i) \times \mu_a(i)}$$
(↑ $\mu_g(i)$?)

con $C_j^{(i)}$ **BLOCCO di JORDAN**

$$C_j^{(i)} = \begin{bmatrix} \lambda_i & 1 & 0 \\ & \ddots & 1 \\ 0 & & \lambda_i \end{bmatrix} \in \mathbb{C}^{\nu_j^{(i)} \times \nu_j^{(i)}}$$

con $\sum_{j=1}^{\mu_g(i)} \nu_j^{(i)} = \mu_a(i)$

1.3 FORMA CANONICA di SCHUR

TEO: $A \in \mathbb{C}^{n\times n}$ con autovalori $\lambda_1, \dots, \lambda_n$

$\Rightarrow \exists\, U, T \in \mathbb{C}^{n\times n}$ U unitaria ($UU^H = U^H U = I$);
T triang. sup t.c. $\text{diag}(T) = (\lambda_1, \dots, \lambda_n)$

t.c. $A = UTU^H$

dim: Per induzione

- $n = 1$
  Prendo $U = [1]$ ed è ovvio
- $n > 1$
  Considero x_1 autovettore normalizzato (in norma 2) rsp a λ_1 autoval
  $S := \langle x_1 \rangle$
  Considero $S^\perp$ con base ortonormale $(y_2, \dots, y_n)$
  Costruiamo $Q = [x_1 \mid y_2 \mid \dots \mid y_n]$
  Abbiamo che Q è unitaria e $Q^H x_1 = e_1$
  $Q^H Q = \begin{bmatrix} x_1^H x_1 & x_1^H y_2 & \cdots & x_1^H y_n \\ x_1^H y_2 & y_2^H y_2 & & \\ \vdots & & & \\ x_1^H y_n & & & y_n^H y_n \end{bmatrix}$
  $x_1^H x_1 = 1$ perché x_1 è normalizzato
  $= 1$ perché $(y_2, \dots, y_n)$ è ortonormale
  $= 0$ perché $y_2, \dots, y_n$ è base di $S^\perp$

  Poniamo $B = Q^H A Q$

  $Be_1 = Q^H A Q e_1 \overset{(*)}{=} Q^H A x_1 = Q^H \lambda_1 x_1 \overset{(*)}{=} \lambda_1 Q^H x_1 = \lambda_1 e_1$
  (prima colonna di B)

  $$\Rightarrow B = \left[\begin{array}{c|c} \lambda_1 & c^h \\ \hline 0 & A_1 \end{array}\right]$$

Per hp di induzione possiamo scrivere $A_1 = U_1 A_2 U_1^H$ dove U_1 è unitaria e A_2 è triang. sup con autoval $(\lambda_2, \dots, \lambda_n)$

Abbiamo:

$$A = Q B Q^H = Q \left[\begin{array}{c|c} \lambda_1 & \underline{c}^H \\ \hline \underline{0} & U_1 A_2 U_1^H \end{array}\right] Q^H$$

$$= Q \underbrace{\left[\begin{array}{c|c} 1 & \underline{0}^H \\ \hline \underline{0} & U_1 \end{array}\right]}_{\text{prodotto di unitarie è unitario!}} \underbrace{\left[\begin{array}{c|c} \lambda_1 & \underline{c}^H U_1 \\ \hline \underline{0} & A_2 \end{array}\right]}_{= T} \underbrace{\left[\begin{array}{c|c} 1 & \underline{0}^H \\ \hline \underline{0} & U_1^H \end{array}\right]}_{= U^H} Q^H$$

(è ancora unitaria)

TEO: $A \in \mathbb{C}^{n \times n}$ hermitiana, con autoval $\lambda_1, \dots, \lambda_n$

$\Rightarrow \exists\, U$ unitaria t.c. $A = U \Lambda U^H$ $\quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$
e colonne di U base autovettori ortonormale

dim: $A = U T U^H$ per il teo sopra

$$T = U^H A U \rightarrow T^H = U^H A^H U = U^H A U = T$$
($A^H = A$ perché A è Hermit.)

Ma $T = T^H$ + T triangolare $\Rightarrow$ T è diagonale e gli autovalori sono reali

$\Rightarrow$ posso scrivere A come nella Tesi e segue subito che le colonne di U sono base di autovettori (ortonorm perché unitaria) ■

1.4 RAGGIO SPETTRALE

def: $S(A) = \max_i |\lambda_i(A)|$

OSS: Il raggio spettrale NON è una norma

dim: $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ $\quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ $\quad A+B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$S(A) = 0$ $\quad S(B) = 0$ $\quad S(A+B) = 1$

$$S(A+B) > S(A) + S(B)$$ ■

TEO: $\forall \|\cdot\|$ norma matriciale vale $S(A) \leq \|A\|$

dim: Sia λ autoval A che è t.c. $|\lambda| = S(A)$ e v autovett corrisp.

$V = [\underline{v} \,|\, \dots \,|\, \underline{v}]$

Quindi $AV = \lambda V$. (*)

Abbiamo:

$$|\lambda| \|V\| = \|\lambda V\| \overset{(*)}{=} \|AV\| \leq \|A\| \cdot \|V\| \quad \text{(per proprietà submoltiplic.)} \quad \Rightarrow |\lambda| \leq \|A\|$$ perché $\|V\| \neq 0$ ■

TEO: $A \in \mathbb{C}^{n \times n} \Rightarrow \forall \varepsilon > 0 \;\; \exists \|\cdot\|$ n.m. indotta t.c.

$$S(A) \leq \|A\| \leq S(A) + \varepsilon$$

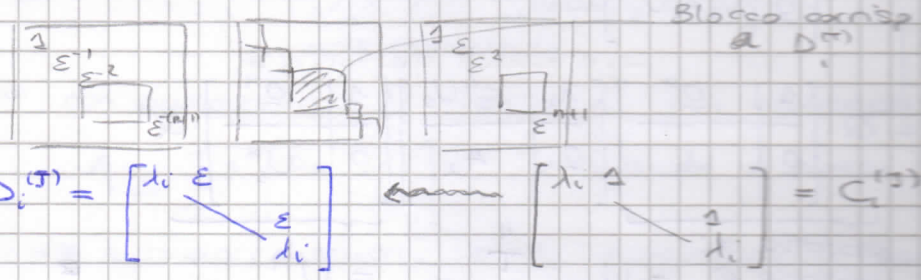
dim: Considero la forma di Jordan $\rightarrow$ $A = T J T^{-1}$

J diag. a blocchi, con blocchi del tipo $C_i^{(j)} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$

Definiamo $E := \text{diag}(1, \varepsilon, \dots, \varepsilon^{n-1})$

$E^{-1} J E = E^{-1}(T^{-1}AT)E$ è diagonale a blocchi
(perché è diag · diag blocchi · diag)

Come sono fatti i blocchi?



$$D_i^{(T)} = \begin{bmatrix} \lambda_i & \varepsilon & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon \\ & & & \lambda_i \end{bmatrix} \quad \text{circa} \quad \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} = C_i^{(T)}$$

$$\|E^{-1}JE\|_\infty = \max_{i,j} \|D_i^{(T)}\|_\infty \overset{(**)}{=} \max_i (|\lambda_i| + \varepsilon) \le S(A) + \varepsilon$$

↑ il max sulla matrice è uguale al max delle norme dei blocchi perché quando faccio il max su riga (colonna) le "righe non si intersecano"

Ma $\|E^{-1}JE\|_\infty = \|E^{-1}(T^{-1}AT)E\|_\infty = \|A\|_*$ norma matriciale indotta

PROP: $\|\cdot\|_* : \mathbb{C}^{n\times n} \longrightarrow \mathbb{R}^+ \cup \{0\}$ t.c. $\|A\|_* = \|S^{-1}AS\|_\infty$ è norma matriciale indotta.

dim: (traccia)

- $\|x\|_* = \|S^{-1}x\|_\infty$ è norma vettoriale
- Si considera la n. m. indotta da $\|\ \|_*$ e si verifica che è quella dell'enunciato.

CONSEGUENZA: $S(A) = \inf_{\|\cdot\| \text{ n. m. i.}} \|A\|$ ma in generale non è un min

OSS: Se gli autoval. che realizzano $S(A)$ hanno $\mu_g = \mu_a$

$\Rightarrow \exists \|\cdot\|$ n.m. indotta t.c. $S(A) = \|A\|_*$

"dim": Qui (**) per il raggio spettr. (autoval max) posso non mettere $+\varepsilon$ perché $D_i^{(T)}$ è diag dato che $\mu_g = \mu_a$ ☐

Questa condizione vale sicuramente se A è diagonalizzabile.
Inoltre se A è Hermitiana so che $\|\cdot\|_* = \|\cdot\|_2$

1.5 SISTEMI LINEARI

PROBLEMA: $Ax = b$ $A \in \mathbb{C}^{n\times n}$ non singolare, $x, b \in \mathbb{C}^n$

TEO: $Ax = b$ $A \in \mathbb{C}^{m\times n}$ $b \in \mathbb{C}^m$ $x \in \mathbb{C}^n$
consistente sse $\text{rango}(A|b) = \text{rango}(A)$

Possiamo formalizzare il nostro problema così:

$f: \mathbb{R}^N \longrightarrow \mathbb{R}$, valutare f in x assegnato

Abbiamo 3 tipi di errore:

- ERRORE INERENTE ⟹ condizionamento P_b
- ERRORE ALGORITMICO ⟹ stabilità metodo/algoritmo
- ERRORE ANALITICO ⟹ convergenza

L'errore inerente è generato dagli errori di rappresentazione dei dati. Devo vedere se questi errori di rappresentazione vengono amplificati! (se sì MAL condizionato) $x \rightsquigarrow fl(x)$

L'errore algoritmico è generato da errori nelle operazioni in aritmetico floating point (N.B visto che l'algoritmo lo posso cambiare devo usare metodi stabili) $fl(x)+fl(y)$ è ? MJ se no devo forse ie float

L'errore analitico ce l'ho ad esempio quando passo da un problema continuo a uno discreto, o quando invece di fare un limite faccio un troncamento.

Quindi esaminiamo bene ie nostro problema:

$$\left.\begin{array}{l} x \overset{(IN)}{\rightsquigarrow} fl(x) \\ f \overset{(AN)}{\rightsquigarrow} g \rightsquigarrow \psi \end{array}\right\} \begin{array}{l} \text{? fl approssimata se facessi bene i conti} \\ f(x) \rightsquigarrow \psi(fl(x)) \\ \downarrow \text{valore da calcolare} \qquad \downarrow \text{valore calcolato} \end{array}$$

$$\Rightarrow E_{tot} = \frac{\psi(fl(x)) - f(x)}{f(x)} \overset{(teo)}{=} E_{IN} + E_{ALG} + E_{AN}$$

Quindi posso considerare i 3 errori separatamente per avere l'errore totale.

$$E_{AN} = \frac{g(x) - f(x)}{f(x)}$$

$$E_{IN} = \frac{f(fl(x)) - f(x)}{f(x)}$$

$$E_{ALG} = \frac{\psi(fl(x)) - g(fl(x))}{g(fl(x))}$$ ← Ci vede solo g e fl(a), non sa neanche che c'è una f e una x

(*) Se f è sufficientemente regolare

$$E_{IN} = \sum_{i=1}^{n} c_i \varepsilon_{x_i}$$

con $\varepsilon_{x_i} = \frac{fl(x_i) - x_i}{x_i}$, $|\varepsilon_{x_i}| \leq u \ \forall i$ e $c_i = \frac{x_i}{f(x)} \cdot \frac{\partial f}{\partial x_i}$ COEFFICIENTI di AMPLIFICAZIONE

23/10/2012 (2)

Sia ora $G: \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$G(\tilde{b}) = G(b) + J_G(b)\,\delta b$ con $\tilde{b} = b + \delta b$

Abbiamo:

$$\|Er\| = \frac{\|G(\tilde{b}) - G(b)\|}{\|G(b)\|} \leq \frac{\|J_G(b)\|\,\|\delta b\|}{\|G(b)\|} = \frac{\|J_G(b)\|\,\|b\|}{\|G(b)\|} \cdot \frac{\|\delta b\|}{\|b\|}$$

Ricordiamo che noi vogliamo risolvere $Ax = b$ quindi $x = A^{-1}b$ quindi nel nostro caso

$$G(b) = A^{-1}b$$

Quindi:

$$\frac{\|J_G(b)\|\,\|b\|}{\|G(b)\|} = \frac{\|A^{-1}\|\,\|b\|}{\|A^{-1}b\|} = \frac{\|A^{-1}\|\,\|Ax\|}{\|x\|} \leq \frac{\|A^{-1}\|\,\|A\|\,\|x\|}{\|x\|} = K(A)$$ NUMERO di CONDIZIONAM.

In conclusione, chiamato $\varepsilon_x = \frac{\|\delta x\|}{\|x\|}$, abbiamo ottenuto che

$$\varepsilon_x \leq K(A) \cdot \varepsilon_b$$

quindi $K(A)$ rappresenta di quanto viene amplificato l'errore.

def: Il **NUMERO di CONDIZIONAMENTO** è $K(A) = \|A\| \cdot \|A^{-1}\|$
Per convenzione $K(A) = +\infty$ se A è singolare.

OSS: $K(A) \geq 1$

Infatti: $K(A) = \|A\| \cdot \|A^{-1}\| \geq \|AA^{-1}\| = \|I\| \geq 1$ (*)

(*) perché qualsiasi sia la norma vale:

$\|I\| = \|I \cdot I\| \leq \|I\| \cdot \|I\| \Rightarrow 1 \leq \|I\|$

def: **CONDIZIONAMENTO SPETTRALE**. $K_2(A) = \|A\|_2 \cdot \|A^{-1}\|_2$

Ricordiamo che abbiamo visto che

$$\|A\|_2 = S(A^H A)^{1/2} = \sqrt{\lambda_{max}(A^H A)}$$
(hermitiana)

La matrice $A^H A$ è simmetrica, semidef. positiva, infatti:

$\forall x \neq 0 \quad x^H A^H A x = \underset{y = Ax}{y^H y} = \|y\|_2^2 \geq 0$

($x \neq 0$ e A non singolare $\Rightarrow y \neq 0$ dove è anche def. pos.)

$$\|A^{-1}\|_2 = S^{1/2}(A^{-H}A^{-1}) = S^{1/2}(AA^H)^{-1}$$
$$\overset{!}{=} \sqrt{\max_i \left(\frac{1}{\lambda_i(AA^H)}\right)} = \sqrt{\frac{1}{\lambda_{min}(AA^H)}} = \sqrt{\frac{1}{\lambda_{min}(A^H A)}}$$
(perché $AA^H \sim A^H A$)

Quindi alla fine otteniamo:

$$K_2(A) = \sqrt{\frac{\lambda_{max}(A^H A)}{\lambda_{min}(A^H A)}}$$ se A è NON singolare

Se inoltre A è hermitiana $\|A\|_2 = S(A)$ e quindi:

$$K_2(A) = \frac{\max_{i=1,\dots,n} |\lambda_i(A)|}{\min_{i=1,\dots,n} |\lambda_i(A)|} \qquad \lambda_i(A) \in \mathbb{R}$$ (perché A hermit.)

Se A è anche hermitiana definita positiva

$$K_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} \qquad \lambda_i(A) \in \mathbb{R}^+$$

1.6 STIMA di $K(A)$

$\forall \|\cdot\|$ n.m. (indotta) $\quad S(A) \leq \|A\|$

$$\Rightarrow K(A) = \|A\| \cdot \|A^{-1}\| \geq S(A) \cdot S(A^{-1}) = \frac{\max |\lambda_i(A)|}{\min |\lambda_i(A)|}$$

Quindi anche usando una norma generica abbiamo una stima di $K(A)$ che coinvolge solo gli autovalori di A.

Questa stima è dal basso, quindi ci dice quanto è almeno mal condizionato il problema.

1.7 METODI di FATTORIZZAZIONE

Il nostro problema è sempre risolvere $Ax = b$ con $A \in \mathbb{C}^{n \times n}$, $x, b \in \mathbb{C}^n$, A non singolare.

Abbiamo dei casi banali:

- $A = D$ diagonale $\rightarrow x_i = b_i / a_{ii}$
- $A = L$ triang. inf $\rightarrow$ ris. forward
- $A = U$ $\rightarrow$ ris. backward $\quad x_i = \dfrac{(b_i - \sum_{j=i} a_{ij} x_j)}{a_{ii}}$

Metodi diretti $\rightsquigarrow$ fattorizzazione di A

$$A = BC$$

$$\Rightarrow BCx = b$$

pongo $y = Cx \rightarrow$ Risolvo (I) $By = b$

(II) $Cx = y$

Chiaramente B e C devono essere matrici comode.

Abbiamo vari metodi di fattorizzazione:

- LU con $\text{diag}(L) = I$ Gauss $n^3/3$
- $LL^H = R^H R$ con $\text{diag}(L), \text{diag}(R) > 0$ Cholesky [solo se A è hermit. def. positiva] $n^3/6$

(Questi due metodi sono praticamente lo stesso cosa se A è definita positiva, quale il costo comp. è lo stesso in quel caso.)

- QR con Q unitaria e R triang. sup. $2n^3/3$

OSS: Risolvere $Qy = b$ è comodo lo stesso perché equivale a

$$y = Q^H b$$ dato che Q è unitaria

(Il costo è il doppio di Gauss dato che invece di risolvere, cioè un sistema triangolare devo fare un prodotto matrice vettore)

OSS: Se A è reale allora anche le matrici delle fattorizzazioni lo sono

OSS: LU non è detto che lo possa sempre fare, ma se c'è è unica.
QR c'è sempre ma non è detto sia unica, ma questo cosa può fare avere comodo.

1.8 FATTORIZZAZIONE LU

A matrice da fattorizzare $\rightarrow A^{(1)} = A$

Pongo $A^{(k+1)} = M_k A^{(k)}$ con M_k matrice elementare di Gauss al passo k $\quad k = 1, \dots, n-1$

$$M_k = \left[\begin{array}{c|c} I_{k-1} & 0 \\ \hline 0 & \begin{array}{cc} 1 & \\ -m_{ik} & \ddots \end{array} \end{array}\right] \qquad \text{con} \quad m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$$

Ma possiamo scrivere M_k anche nella forma:

$$M_k = I - \underline{m}_k \cdot \underline{e}_k^T \qquad \text{con} \quad \underline{m}_k = \begin{bmatrix} 0 \\ \hline m_{ik} \\ \vdots \end{bmatrix} \} k$$

$\uparrow$ DIADE $|\cdot|$ = matrice

Abbiamo:

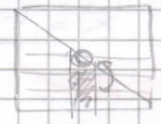
$$(M_{n-1} \cdot \dots \cdot M_1) A = U$$

$$(M_{n-1} \cdot \dots \cdot M_1)^{-1} = M_1^{-1} \cdot \dots \cdot M_{n-1}^{-1} = L$$

TEO: $A \in \mathbb{C}^{n\times n}$. Se A_k sottomatrice principale di testa è non singolare $\forall\, k = 1,\dots,n$ $\exists!$ $A = LU$

TEO: $A \in \mathbb{C}^{n\times n}$. $\exists$ P matrice di permutazione t.c. $PA = LU$

N.B.: Nel metodo in pratica si usa sempre la permutazione perché così viene meglio per i calcoli.
(Pivot parziale, Pivot totale)
→ EFFETTO STABILIZZANTE:

Pivot parziale: 

→ 1. $|m_{ik}| \le 1$ (dato che $a_{ik} \le a_{kk}$ visto che metto il max in a_{kk})

→ 2. $a_M^{(n)} \le 2^{n-1} a_M^{(1)}$

Pivot totale in pratica non si usa quasi mai, perché è meglio ma non serve di solito.

ESEMPIO: $A = \begin{bmatrix} \varepsilon & 1 \\ 1 & 0 \end{bmatrix}$ $\quad b = \begin{bmatrix} 1+\varepsilon \\ 1 \end{bmatrix}$ $\quad \varepsilon > 0$ $\quad x_{esatta} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Il problema è ben condizionato: $A^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -\varepsilon \end{bmatrix}$

$K_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = (1+\varepsilon)^2 \approx 1$ per ε piccolo

Però se faccio LU ottengo:

$$A = LU = \begin{bmatrix} 1 & 0 \\ 1/\varepsilon & 1 \end{bmatrix} \begin{bmatrix} \varepsilon & 1 \\ 0 & -1/\varepsilon \end{bmatrix}$$

Per ε piccolo esplodono !!!

→ bisogna fare il pivot parziale!

Ad esempio si può verificare che in $\mathbb{F}(10,3)$ (aritmetica di macchina) con $\varepsilon = 0{,}3 \cdot 10^{-3}$ si ottiene $\tilde{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

1.9 FATTORIZZAZIONE QR

def: **MATRICE ELEMENTARE di HOUSEHOLDER:**

$$P = I - \beta\, \underline{v} \cdot \underline{v}^H \qquad \beta \in \mathbb{R},\ \underline{v} \in \mathbb{C}^n$$

PROP: P è hermitiana.
P è unitaria se $\beta = \dfrac{2}{\|\underline{v}\|_2^2}$

(i.e. dato un vettore $\underline{v}$ posso associargli un'unica matrice unitaria e herm. in questa forma)

dim: $P^H = I - \beta (\underline{v}^H)^H \cdot \underline{v}^H = I - \beta \underline{v}\,\underline{v}^H = P$ ok

Deve valere $I = PP^H = (I - \beta \underline{v}\underline{v}^H)(I - \beta \underline{v}\underline{v}^H)$ ($P = P^H$)

$$= I - 2\beta \underline{v}\underline{v}^H + \beta^2 \underline{v}\underline{v}^H \underline{v}\underline{v}^H$$

$$= I - 2\beta \underline{v}\underline{v}^H + \beta^2 \|\underline{v}\|_2^2\, \underline{v}\underline{v}^H$$

Questo vale $\iff$ $-2\beta \underline{v}\underline{v}^H + \beta^2 \|\underline{v}\|_2^2\, \underline{v}\underline{v}^H = 0$

$$(-2 + \beta \|\underline{v}\|_2^2)\, \beta \underline{v}\underline{v}^H = 0$$

Se $\beta \neq 0$ (caso interessante) deve essere $-2 + \beta \|\underline{v}\|_2^2 = 0$

→ P relativa al vettore $\underline{v}$, non P·v ! 😉

PROP: Sia $P = P(\underline{v}) = I - \beta \underline{v}\underline{v}^H$ con $\beta = 2\|\underline{v}\|_2^{-2}$. Allora:

- $\forall \underline{x} \in \underline{v}^\perp = \{\underline{x} : \underline{x}^H \underline{v} = 0\}$ vale $P\underline{x} = \underline{x}$ (i.e. P è identità su $\underline{v}^\perp$)
- $P(\underline{v})\gamma\underline{v} = -\gamma\underline{v}$ (i.e. è una riflessione nello stesso da $\underline{v}$)

dim: $P\underline{x} = (I - \beta \underline{v}\underline{v}^H)\underline{x} = I\underline{x} - \beta \underline{v}\underbrace{\underline{v}^H\underline{x}}_{=0} = \underline{x}$

$P(\underline{v})\gamma\underline{v} = (I - \beta\underline{v}\underline{v}^H)\gamma\underline{v} = \gamma\underline{v} - \beta\gamma\|\underline{v}\|_2^2 \cdot \underline{v}$ con $\beta = \frac{2}{\|\underline{v}\|_2^2}$

$= \gamma\underline{v} - 2\gamma\underline{v} = -\gamma\underline{v}$ ■

24/10/2012 ④

TEO: $\forall \underline{x} \in \mathbb{C}^n$, $\{\neq 0\}$ $\exists P \in \mathbb{C}^{n\times n}$ m. e. di Householder t.c.

$P\underline{x} = \alpha \underline{e}_1$ $\alpha \in \mathbb{C}$ opportuno, $\underline{e}_1 = [1, 0, \dots 0]^t$

dim: (I) troviamo $\alpha \in \mathbb{C}$
(II) troviamo $\underline{v} \in \mathbb{C}^n$ che individua P

(I) Sia $\alpha = |\alpha| e^{i\theta}$ (lo scrivo in forma exp)

Vogliamo che $P\underline{x} = \alpha\underline{e}_1$ e quindi deve essere

$\|P\underline{x}\|_2 = \|\alpha \underline{e}_1\|_2 = |\alpha|$

Ma P è unitaria e quindi $\|P\underline{x}\|_2 = \|\underline{x}\|_2 \Rightarrow |\alpha| = \|\underline{x}\|_2$ (*)

Deve anche essere $\underline{x}^H P\underline{x} = \underline{x}^H \alpha \underline{e}_1$ (ovviamente visto che $P\underline{x} = \alpha\underline{e}_1$)

Abbiamo $\underline{x}^H \alpha \underline{e}_1 = \alpha \underbrace{\underline{x}^H \underline{e}_1}_{\bar{x}_1} \in \mathbb{R}$ (**) perché P è hermitiana

$\underline{y}^H P \underline{y} = (\underline{y}^H P \underline{y})^H = \underline{y}^H P^H \underline{y} = \underline{y}^H P \underline{y} \in \mathbb{R}$ ($P = P^H$)

$x_1 = |x_1| e^{i\varphi} \neq 0$ (***)

Abbiamo:

$\alpha \bar{x}_1 = |\alpha| e^{i\theta} |x_1| e^{-i\varphi} \overset{(*)}{=} \|\underline{x}\|_2 |x_1| e^{i(\theta-\varphi)} \in \mathbb{R}$ (**)

$\Rightarrow \theta - \varphi \in k\pi$ e quindi $\theta = \varphi + \delta\pi$ $\delta \in \{0, 1\}$

e quindi:

$\alpha = |\alpha| e^{i\theta} = \|\underline{x}\|_2 e^{i(\varphi+\delta\pi)} = \pm \|\underline{x}\|_2 e^{i\varphi} = \pm \|\underline{x}\|_2 \frac{x_1}{|x_1|}$ (nella pratica uso il segno −)

Se $x_1 = 0$ per convenzione si pone $\alpha = -\|\underline{x}\|_2$ (ho libertà di scelta per θ) in modo che sia soddisfatta (*).

(II) $P\underline{x} = (I - \beta\underline{v}\underline{v}^H)\underline{x} = \underline{x} - \underbrace{\frac{2}{\|\underline{v}\|_2^2} \cdot (\underline{v}^H\underline{x})}_{:=\hat{\beta}=\hat{\beta}(\underline{v})} \cdot \underline{v}$ ($(\underline{v}^H\underline{x})$ è un prodotto scalare quindi è un numero e posso spostarlo davanti)

Si vuole $\hat{\beta}\underline{v} = \underline{x} - P\underline{x} = \underline{x} - \alpha\underline{e}_1$ (***)

Questa relazione non è tanto bella perché $\hat{\beta}$ dipende da $\underline{v}$ e quindi la relazione non è lineare! (...)

OSS: $P[\gamma v] := I - \frac{2}{\|\gamma\underline{v}\|_2^2}(\gamma\underline{v})(\gamma\underline{v})^H = I - \frac{2}{|\gamma|^2\|\underline{v}\|_2^2} |\gamma|^2 \underline{v}\cdot\underline{v}^H = P[\underline{v}]$

quindi la matrice generata da $\underline{v}$ e da $\gamma\underline{v}$ è la stessa!!

(...) Quindi noi possiamo prendere $\hat{\beta} = 1$ tanto la scalatura viene mangiata (basta che sia solo $\hat{\beta} \neq 0$)

$\Rightarrow$ prendo $\underline{v} = \underline{x} - \alpha\underline{e}_1$

Ora ie punto e': posso davvero escludere che $\hat{\beta}=0$?

Per def $\hat{\beta}=0$ sse $v^H x = 0$ sse $x \in V^\perp$ (*)

Dato che x ce l'ho e v lo sto scegliendo basta che lo scelgo in modo che (*) non valga!

Osserviamo che la nostra scelta va bene:

$$v^H x = (x - \alpha e_1)^H x = x^H x - \bar{\alpha} e_1^H x$$
$$\overset{!}{=} \|x\|_2^2 - \bar{\alpha} x_1 = \|x\|_2^2 + \|x\|_2^2 \frac{\bar{x}_1}{|x_1|} \cdot x_1 \qquad \bar{x}_1 x_1 = |x_1|^2$$
$$\overset{!}{=} \|x\|_2^2 + |x_1| \|x\|_2^2 \neq 0 \quad (>0)$$

Ricapitolando:

$$\alpha = \begin{cases} \pm \|x\|_2 \frac{x_1}{|x_1|} & \text{se } x_1 \neq 0 \\ -\|x\|_2 & \text{se } x_1 = 0 \end{cases}$$

$$v = x - \alpha e_1 = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Cos'è v_1? Prendiamo $\alpha = -\|x\|_2 \frac{x_1}{|x_1|}$

$$v_1 = x_1 - \alpha = x_1 - \left(-\|x\|_2 \frac{x_1}{|x_1|}\right) = x_1 + \frac{x_1}{|x_1|}\|x\|_2 = \frac{x_1}{|x_1|}(|x_1| + \|x\|_2)$$

quindi scelgo α di segno "$-$" per questo ← NON RISCHIO di AVERE ERRORI di CANCELLAZIONE

OSS: Se $x \in \mathbb{R}^n \to \alpha \in \mathbb{R}$ e $v \in \mathbb{R}^n$ per costruzione
$\Rightarrow P$ è ortogonale e $P \in \mathbb{R}^{n \times n}$

Ricordiamo che volevamo scrivere $A = QR$. Allora cosa facciamo in generale al passo k-esimo. (Abbiamo visto quello primo)

Passo k-esimo: a_k sottocolonna k-esima $\in \mathbb{C}^{n-k}$

$v_k \in \mathbb{C}^n$, $v_k = \begin{bmatrix} 0 \\ v_{kk} \\ a_{k+1,k} \\ \vdots \\ a_{nk} \end{bmatrix} \} k-1$

Quindi: procedo facendo $P^{(n-1)} \cdot \ldots \cdot P^{(2)} \cdot P^{(1)} A = \begin{bmatrix} \times & \times & \times \\ & \ddots & \times \\ 0 & & \times \end{bmatrix} = R$

Ma allora abbiamo:

$$A = (P^{(n-1)} \cdot \ldots \cdot P^{(1)})^{-1} R$$
$$\overset{!}{=} P^{(1)-1} \cdot \ldots \cdot P^{(n-1)-1} R = P^{(1)} \cdot \ldots \cdot P^{(n-1)} R = QR$$

perché $P^{-1} = P^H = P$ ← prodotto di matrici unitarie è unitaria

Vediamo ora la NON UNICITÀ della scomposizione QR.

$S = \text{diag}(\theta_1, \ldots, \theta_n)$ unitaria $|\theta_i| = 1 \ \forall i$ MATRICE di FASE

$$A = QR = QSS^H R = (QS)(S^H R) = \tilde{Q}\tilde{R}$$

$\tilde{Q}$ è unitaria perché prodotto di unitarie.

$\tilde{R}$ è triangolare sup perché S^H è diagonale.

Quindi ho infiniti modi per fattorizzare col metodo QR.

Abbiamo che il costo del metodo QR è $O(n^3)$ ed è doppio rispetto a quello LU (è sempre $O(n^3)$ ma doppio)

Dato che il costo è doppio di solito si usa il metodo LU.
La fattoriz. QR si usa però per esempio per il calcolo del rango di una matrice o all'interno del metodo svd...

Esiste una tecnica di pivot anche per il metodo QR.
A che cosa serve? (Visto che non la fattorizzazione esiste)
Generalmente serve per dare un ordinamento negli elementi della diagonale della fatt.

MASSIMO PIVOT per COLONNE in QR:

Passo k $\quad A^{(k)} = \begin{bmatrix} \mathbb{I}_{k-1} & \\ 0 & \tilde{A}^{(k)} \end{bmatrix}$ $\quad k-1$

cerco la colonna di $\tilde{A}^{(k)}$ con norma 2 max e scambiare le colonne. ($\Pi^{(k)}$ matrice di scambio)

Alla fine ho $\quad P^{(n-1)} \cdot \dots \cdot P^{(1)} A \underbrace{\Pi^{(1)} \cdot \dots \cdot \Pi^{(n-1)}}_{=\Pi} = R$ (moltiplico da qui perché il lavoro sulle colonne)

$\Rightarrow A\Pi = QR$

(*) In questo modo la matrice R è t.c. $|r_{11}| \geq |r_{22}| \geq \dots \geq |r_{nn}|$

dim: Nella notazione del TEO di esistenza di $P^{(i)}$ abbiamo

$r_{ii} = \alpha_i$

Avevamo visto che $|\alpha_1| = \|\underline{a}_1\|_2$ con $A\Pi^{(1)} = [\underline{a}_1 | \underline{a}_2 | \dots | \underline{a}_n]$
per costruzione di $P^{(1)}$

E' quindi evidente che $|\alpha_1| = \|\underline{a}_1\|_2 \geq \|\underline{a}_i\|_2 \quad i = 2, \dots, n$ perché ho scelto $\Pi^{(1)}$

So anche però che $\|\underline{a}_i\|_2 = \|P^{(1)}\underline{a}_i\|_2$ dato che $P^{(1)}$ è unitaria.

Chiaramente $A^{(2)} = P^{(1)} A \Pi^{(1)} \Pi^{(2)}$

$A^{(2)} = \begin{bmatrix} * & * \\ 0 & \tilde{A}^{(2)} \end{bmatrix}$

Chiaramente $\|P^{(1)}\underline{a}_i\|_2 \geq \|\tilde{\underline{a}}_i\|_2$ perché
$\tilde{\underline{a}}_i$ è uguale a $P^{(1)}\underline{a}_i$ meno delle componenti in meno

Quindi alla fine avrò che $|\alpha_1| \geq \|\tilde{\underline{a}}_2\|_2 = |\alpha_2|$

poi si itera per gli altri α_i. Dato che $\alpha_i = r_{ii}$ ottengo la tesi ■

QR per calcolo del RANGO
Vediamo come usiamo questo metodo per calcolare il rango.

$A \in \mathbb{C}^{n \times n}$. Se $rk(A) = p < n \Rightarrow$ applico QR con pivot max

dopo p passi si ferma $A\Pi = QR$ $\quad \det(R_1) \neq 0$

$R = \begin{bmatrix} R_1 & S \\ 0 & 0 \end{bmatrix} \begin{matrix} \} p \\ \} n-p \end{matrix}$ $\quad p \quad n-p$

FATTORIZZAZIONE QR PER MATRICI RETTANGOLARI

$A \in \mathbb{C}^{m \times n} \quad m > n \quad A = QR$

$\begin{matrix} m \\ \end{matrix} \begin{bmatrix} A \end{bmatrix} = \begin{matrix} m \\ \end{matrix} \begin{bmatrix} Q \\ \text{unitaria} \end{bmatrix} \cdot \begin{matrix} m \\ \end{matrix} \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix}$

$n \qquad m \qquad n$

$\tilde{R}$ triang. superiore $\in \mathbb{C}^{n \times n}$

OSS: Il procedimento è lo stesso cambia solo che ho n passi perché devo azzerare anche l'ultima colonna. (Se è quadrata non ho niente da azzerare nell'ultima colonna)

$\text{COSTO} = n^2\left(m - \frac{n}{3}\right)$